EdSource®

*Clarifying
Complex
Education
Issues*

Lining Up Public Schools:
California's Academic Performance Index (API)

A complex calculation called the Academic Performance Index is currently the means by which California is publicly ranking its schools. How is the API computed? Is it a fair measurement of a school's performance? And will this index help the state improve the quality of its schools?

In January 2000, California embarked upon a dramatically different approach to focusing schools' attention on student performance. For the first time, schools are to be publicly ranked using a complex formula—the Academic Performance Index (API)—which tracks and compares academic performance and improvement annually.

Underscoring the importance of this new school quality measure, Governor Gray Davis is calling the API the "North Star" of the Public Schools Accountability Act (PSAA), the state's plan to make schools more accountable. Passed by the Legislature in April 1999, the PSAA rewards schools that meet performance goals and prescribes interventions for schools that fall short of these expectations.

This EdSource report explains the basic principles and steps involved in calculating schools' individual API scores, how schools will be ranked, how much they are expected to improve academically, and how their performance compares to other schools. It also summarizes initial reactions from various stakeholders, including educators and state officials, about the promise and potential pitfalls of this new system.

In a nutshell: How the 1999 API will work

In November 1999, the State Board of Education (SBE) unanimously adopted most of an advisory committee's recommendations for establishing the API. Since then, the California Department of Education (CDE) has moved ahead quickly to derive and make public the first round of API information, the "base year 1999 API."

That information includes a 1999 API score for each school, statewide ranks, and ranks based on comparisons to schools with similar characteristics. Also released are 1999 growth targets and 2000 API targets for each school and for each major ethnic and economically disadvantaged student subgroup within a school.

The state chose to release this information nearly six months earlier than the legislatively mandated deadline of June 2000. That way schools could be given advance warning of how much (or how little) they need to improve academically to be eligible for the 1999–2000 PSAA awards program. Likewise, underachieving schools that participate in the PSAA's intervention program can use this information to refine their school improvement plans, which are required as part of the program. Despite this early release, participating schools still face a timing dilemma. According to legislative deadlines, they should already be well along the way to developing their improvement plans, with final plans due to their local school boards by March 15.

Not all public schools will receive an API score or be ranked this year. Schools

that did not test at least 65% of their students, schools with fewer than 100 students, and several other categories of special schools were excluded from this year's API calculations and will have a separate accountability system created for them by this summer.

The 1999 API calculation is based on several principles

The PSAA Advisory Committee—which the Legislature charged with making recommendations on the design of the API—considered many complex issues in developing the formula for arriving at a single API number for a school. Below is a brief explanation of the various factors and reasoning involved.

Only one indicator meets the criteria for inclusion in the first API. The SAT-9 (the Stanford Achievement Test, Ninth Edition, Form T) is the sole indicator for determining a school's performance in the 1999 API. Envisioning a broader approach for measuring school performance, the PSAA calls for an index that relies on multiple measures of school performance. However, by law these measures can only be included if they are found to meet certain criteria of trustworthiness (i.e., producing valid and reliable results).

The PSAA Advisory Committee found that only the SAT-9 portion of the state-mandated Standardized Testing and Reporting (STAR) program met these criteria. The SAT-9 portion is administered each spring to all students in grades 2 to 11. Results for districts and schools are publicized each July.

Additional questions have been added to the SAT-9 in the past year to more precisely assess the success of students in meeting state content standards. Including student scores on these questions was also considered by the committee. However, these standards-based, augmented items still lack the technical accuracy needed for inclusion in the API.
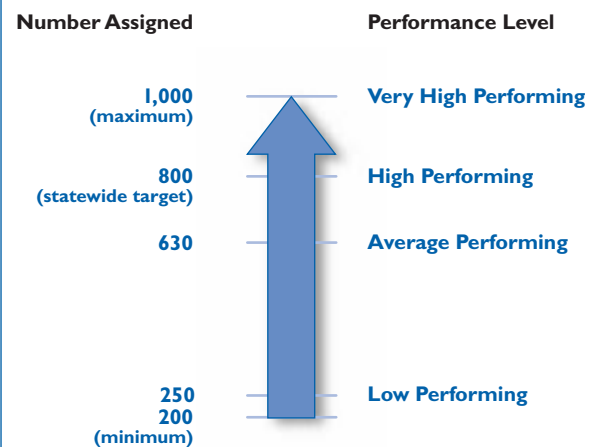
Other performance indicators mentioned by the PSAA law—such as student attendance,

The Three Basic Elements of the Public Schools Accountability Act (PSAA)

1. **The Academic Performance Index**, considered the cornerstone of this new law, produces a single numerical rating of a school's performance. That number serves as the basis for ranking schools, calculating how much they must improve their performance each year, and comparing their growth to similar schools.

2. **The Immediate Intervention/Underperforming Schools Program (II/USP)** provides assistance and intervention to a selected set of schools that do not meet their growth targets.

3. **The Governor's Performance Award Program (GPAP)** provides monetary and non-monetary rewards to schools that meet or exceed performance targets or demonstrate high achievement.

EdSource 1/00

Figure 1
Performance Scale for Schools

Number Assigned	Performance Level
1,000 (maximum)	**Very High Performing**
800 (statewide target)	**High Performing**
630	**Average Performing**
250 **200** (minimum)	**Low Performing**

Schools are expected to aim for 800, the interim statewide performance target.

Data: California Department of Education   EdSource 1/00

school staff attendance, and graduation rates—were either unavailable or were not yet uniformly collected from schools. Further,

Figure 2
Performance Band Weights

SAT-9 Performance Bands	Weights
Band 5: 80th-99th NPR	1,000
Band 4: 60th-79th NPR	875
Band 3: 40th-59th NPR	700
Band 2: 20th-39th NPR	500
Band 1: 1st-19th NPR	200

National percentile ranks (NPR) are sorted by performance bands and then assigned weights.

Data: State Board of Education Adoption, November 1999

EdSource 1/00

the results of two other legislatively mandated tests—a standards-based matrix test and a high school exit exam—were also slated for inclusion in the API. But these assessments have not yet been developed.

A 200-1,000 numerical scale was selected to assign API scores. The PSAA Advisory Committee decided schools should be assigned API scores based on a scale of 200-1,000. No school can generate an API score less than 200. While a scale of 1-100 may seem more intuitive, the committee reasoned that it would be too easily confused with the way in which SAT-9 scores and percentages are currently reported. The SBE adopted this recommendation.

As Figure 1 reveals, a school that scores 1,000 on the API is considered a "very high performing school," whereas a school that scores 250 is considered a "low performing school." The PSAA also calls for a statewide performance target that represents a high level of performance to which all schools should aspire. Until the state adopts performance standards based on recently adopted content standards, the interim statewide performance target for all schools is 800. This number is the measure against which each school's individual growth will be calculated for the 1999–2000 school year. According to the CDE, only 13 percent of all elementary schools, 11 percent of all middle schools, and 5 percent of all high schools are likely to have an API score of 800 or higher.

Nationally normed scores allow California to compare its performance to that of other states. Only nationally normed individual student scores, that is national

Not All STAR Results Are Included in the 1999 API

The following statewide STAR results are not included in this year's API:

✔ **SAT-9 scores of recently enrolled students.** To take into consideration student mobility, the PSAA specifies that all students who have been in the district less than a year will not have their SAT-9 scores calculated as part of the index. This also means that the scores of English learners (ELs) who have been in the district for less than a year are excluded. All other ELs will have their scores count as part of the API.

✔ **Standards-based augmented scores.** Scores from the augmented test items of STAR that are based on the state's content standards were also excluded. Their validity and reliability are still unknown, and the state has yet to adopt performance standards that are critical to scoring these results.

✔ **Primary language scores.** Another required element of STAR is a primary language test. SABE/2 (the Spanish Assessment of Basic Education, Second Edition) is the test given to all Spanish-speaking ELs who have been in the state for less than a year. Districts may also choose to give this test to ELs who have been in the state longer than a year, though they receive no state funds for doing so. Given the inconsistency among districts about which students are given the test and how test results are aggregated at the state level, SABE scores are excluded from the API.

Additional rules about the inclusion/exclusion of test results are explained on the California Department of Education's website, www.cde.ca.gov/psaa/api/base/baseapi.htm.

EdSource 1/00

percentile rank (NPR) scores, are used to calculate the API. The multiple-choice SAT-9 test is nationally normed. This means a national benchmark was set when a sample of 500,000 to 600,000 students who reflect the demographic profile of all U.S. students took the test in 1995. By comparing individual student scores against this benchmark, California can compare its performance to that of the nation.

Performance bands sort student scores and weight improvement over time. A school's API combines NPR student scores in each subject, such as mathematics or reading. The individual subject scores for all students at a school are sorted into five "performance bands" according to their national percentile rank. (See Figure 2.) Four fixed cut

points separate these performance bands. The percentage of students scoring within each of the five performance bands is multiplied by a series of progressive weights. (See explanation on page 5.)

Content area weights are used to focus educators on state curriculum priorities. Additional weights—keyed to content areas (subjects)—are also part of the API equation. (See Figure 3.) These weights vary according to grade level. This occurs because the subjects tested by SAT-9 in grades 2-8 (elementary and middle schools) differ somewhat from the content of the tests used for grades 9-11 (high schools).

For elementary and middle schools, content weights are based on state curriculum priorities set by the SBE. Given the state's emphasis on improving literacy, student scores in reading, language, and spelling at the elementary and middle school level are weighted collectively at 60%, whereas mathematics is weighted at 40%. (See Figure 3.) At the high school level, the PSAA Advisory Committee decided that equal effort should be devoted to all curriculum areas so weights are an even 20% for each content area.

Calculating this year's API score involves several complicated steps
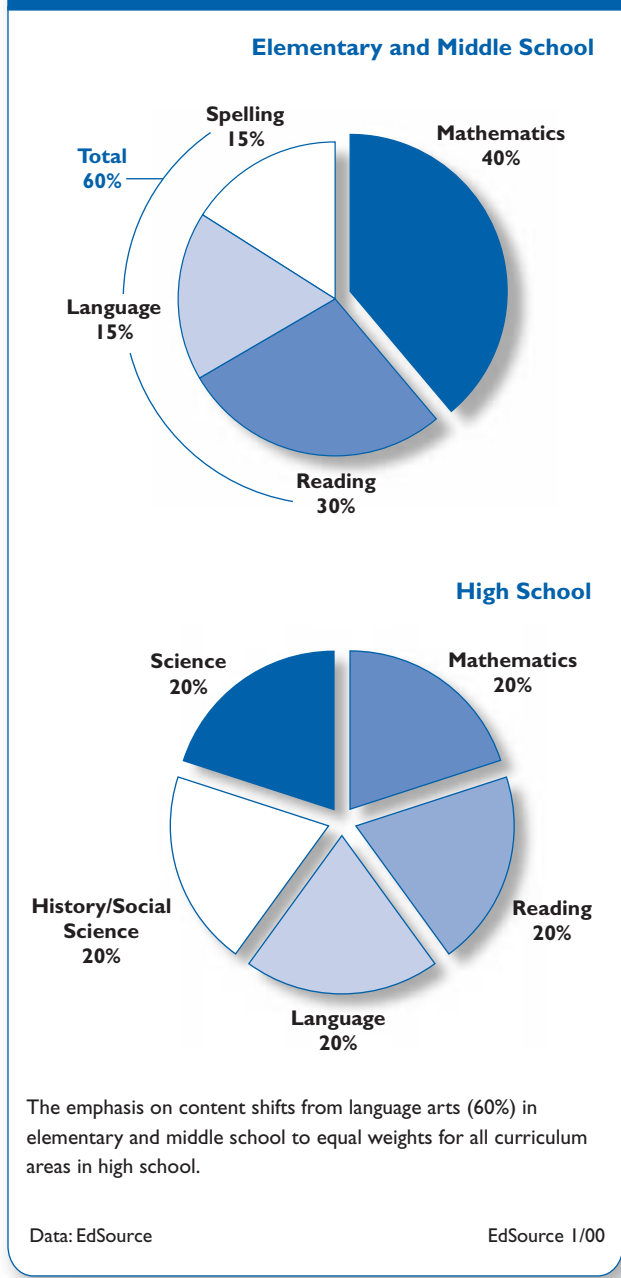
Figures 4 and 5 provide one example of how an elementary school's single API score is derived. This example also shows how performance band and content area weights are applied.

The API equation combines weighted content scores to produce a school's single API score. Figures 4 and 5 show how the weighted total score for one content area is combined with other content area total scores to derive an elementary school's single API score. The calculation in Figure 4 shows that the percentage of students' scores in reading within each of the performance bands (column 3) is multiplied by the appropriate weights (column 2). The end result is a weighted score for each band (column 4). Those scores are summed to yield the total weighted score across all performance bands for a single subject, in this case reading. A similar calculation is done for each content area tested.

A second set of weights is then applied to the content areas (see Figure 5), in this case 30% for reading (column C). The total weighted score for reading is 504 x 30%, or 151 (column D). After the total weighted scores for other content areas are calculated, they are added together to come up with a school's API score for 1999.

API scores serve as the basis for ranking schools. Based on the value of their API score, schools are given a

Figure 3
Content Weights

The emphasis on content shifts from language arts (60%) in elementary and middle school to equal weights for all curriculum areas in high school.

Data: EdSource   EdSource 1/00

"statewide rank." There are separate statewide ranks for each school type—elementary, middle, and high school. Schools given a rank of "1" are assigned to the 1st decile (bottom 10%) and are considered the lowest performing. Schools assigned a rank of "10" are assigned to the 10th decile (top 10%) and are considered the highest performing. In addition to statewide API ranks, schools are assigned "a similar schools rank" based on how their API score compares to other schools with similar demographic characteristics. (See page 6 for further explanation.)

In the future, improvement at the bottom will receive the greatest credit

Once they receive their 1999 base line API scores, schools are supposed to focus their efforts on improving student performance. Driving this focus on improved performance is a progressive system of performance band weights. As Figure 6 illustrates, annual improvements in performance among students in the bottom bands grant more API points than improvements made by students at the top. Hypothetically, students moving from Band 1 to Band 2 in the next year will generate 300 points, whereas students moving from Band 4 to Band 5 will generate only 125 points.

What to Aim For: API Target Year 2000

The PSAA calls for schools to show 5% growth in performance each year. After some deliberation, the PSAA Advisory Committee determined, and the SBE concurred, that a school's growth target should be calculated by taking 5% of the distance between a school's 1999 API and the statewide interim performance target of 800. For example, a school with an API of 700 needs to improve by five points, whereas a school with an API of 300 has a 25-point growth target. Schools that already have an API of 800 or more must maintain such a score in future years.

This growth target plays an important role in programs authorized under the PSAA. To be eligible for the PSAA awards, the law requires schools to show either a 5% improvement annually or meet the statewide performance target. Schools participating in the PSAA interventions program must also show progress toward meeting this 5% growth target to avoid interventions and eventually sanctions.

As seen in the example in Figure 7, the three-step process for determining a school's target number to demonstrate adequate growth is:

Figure 4
Deriving a Total Weighted Score for Reading

Column 1 Performance Bands	Column 2 Performance Band Weights	Column 3 % Student Reading Scores in Each Band	Column 4 Weighted Scores for Each Performance Band
Band 5: 80th–99th NPR	1,000	5%	1,000 x 5% = 50
Band 4: 60th–79th NPR	875	5%	875 x 5% = 44
Band 3: 40th–59th NPR	700	25%	700 x 25% = 175
Band 2: 20th–39th NPR	500	35%	500 x 35% = 175
Band 1: 1st–19th NPR	200	30%	200 x 30% = 60

Total Weighted Score Across Bands for Reading **= 504**

Figure 5
Summing Total Weighted Content Scores to Obtain 1999 API

Column A Content Area	Column B Total Weighted Score Across Bands by Content Area	Column C Content Weights for Elementary School	Column D Total Weighted Score Across Bands for Each Content Area
Reading	504	30%	504 x 30% = 151
Language	588	15%	588 x 15% = 88
Spelling	538	15%	538 x 15% = 81
Mathematics	538	40%	538 x 40% = 215

1999 API Score **= 535**

This hypothetical school would be ranked somewhere between low performing (250) and average performing (630).

Figure 6
Credit Given for Improving Student Performance

SAT-9 Performance Bands	Weights	Point Gain for Movement From Next Lower Band
Band 5: 80th–99th NPR	1,000	125
Band 4: 60th–79th NPR	875	175
Band 3: 40th–59th NPR	700	200
Band 2: 20th–39th NPR	500	300
Band 1: 1st–19th NPR	200	N/A

The state applies progressive weights to encourage schools to focus on improving the scores of the lowest-performing students.

Data: State Board of Education Adoption, November 1999   EdSource 1/00

✔ *Step One:* Subtract a school's 1999 API from the state's target of 800.

✔ *Step Two:* Multiply this number by 5% to obtain the 1999 growth target.

✔ *Step Three:* Add the 1999 growth target to the school's 1999 API to derive the school's API target for the year 2000.

Ethnically diverse and poor students are expected to show "comparable" improvement

Schools must also demonstrate that they are making a concerted effort to ensure certain subgroups of students improve their performance "comparably" to that of other students. Specifically, students who fall into this special category belong to what the PSAA defines as "numerically-significant" ethnic or socioeconomically disadvantaged subgroups.

A *numerically significant subgroup* represents: 1) at least 15% of the total school student population and at least 30 students; or 2) at least 100 students (a criterion added by the PSAA Advisory Committee and approved by the SBE). *Ethnic/racial subgroups* used for calculating the API include African American, not Hispanic; American Indian or Alaska Native; Asian; Filipino; Hispanic or Latino; Pacific Islander; and white, not Hispanic. A *socioeconomically disadvantaged student* is defined as any student who participates in a federally funded free or reduced-price lunch program, or whose parents' education is below the level of high school.

Once numerically significant ethnic and socioeconomic subgroups of students are identified, an API for each is calculated. The same schoolwide API formula is applied to determine the subgroup API.

Growth targets for subgroups, however, are determined differently from the schoolwide growth targets. Subgroups are expected to meet a performance target that is 80% of the schoolwide target—"comparable," but not equal growth. For example, if a school's overall growth target is 13, the subgroup's growth target is 80% of 13, or 10. To be eligible for a PSAA award, this school must show that all of its numerically significant subgroups show at least a 10-point improvement in their API index.

A school's performance is compared to other schools with similar characteristics

Besides the statewide API rank, the PSAA also calls for a second set of rankings that shows how a school's API score lines up next to that of other schools with similar characteristics. This is called a "similar schools rank." The specific characteristics for comparison include but are not limited to:

✔ student mobility,

✔ student ethnicity,

✔ student socioeconomic status,

✔ percentage of teachers fully credentialed,

✔ percentage of teachers who hold emergency credentials,

✔ percentage of pupils who are English learners,

✔ average class size per grade level, and

✔ whether schools operate on a multi-track, year-round schedule.

Figure 7
One Example for Calculating Schoolwide Growth Targets

Step 1

$$800 \text{ (State Target)}$$
$$- 535 \text{ (an example 1999 API Score)}$$
$$= 265$$

Step 2

$$265 \times 5\%$$
$$= 13 \text{ (1999 Growth Target)}$$

Step 3

$$13 + 535$$
$$= 548 \text{ (Year 2000 API Target)}$$

Data: State Board of Education Adoption, November 1999   EdSource 1/00

To carry out this comparison, the SBE adopted a separate calculation to generate the School Characteristic Index (SCI). In short, a school is compared to 100 other schools with similar characteristics. All 100 schools, including the school in question, are then listed in order of their API scores and divided into 10 groups. These groups are ranked from 1 to 10 (highest), and the rank for the school in question is reported. For more information on how the SCI is calculated, see the CDE's *Explanatory Notes for the 1999 Academic Performance Index* posted at www.cde.ca.gov/psaa/.

Pending issues: API heightens attention on performance but is still a work in progress

The development of the API has not been without controversy. Education stakeholders are raising serious questions about the repercussions of this evolving high-stakes accountability system on schools, their staffs, and students. Supporters of this new system view these concerns as an indication that the state is finally on the right track. For the first time in many years, they say, intense debate, scrutiny, and effort are being spent carefully analyzing what it means to measure school and student performance and what it will take to improve that performance.

The test-driven index is raising concerns about the influence on classroom practice

Many state and local officials agree that a sole-indicator API is far from ideal. In fact, even the PSAA Advisory Committee "reluctantly" recommended the API be based only on the SAT-9 after finding other potential indicators to be deficient. In their November 9 proposal to the SBE, the committee cautioned that the "SAT-9 has serious limitations as an accountability instrument for California."

Reflecting the sentiments of many educators throughout the state, the committee also called on the SBE to base the API on "measures linked to California's content and performance standards" as soon as possible. At the core of these concerns is the belief that the test could drive classroom practice in a way that is incongruent with, if not inferior to, the teaching and learning concepts associated with newly adopted state content standards.

Another set of related issues concerns the changing nature of the index's composition. How well can other indicators be incorporated over time without adding confusion for schools about what they should focus on? Given that the

index's indicators are likely to differ each year, will it be a useful measure of the change in school performance over time? Or will it be akin to comparing apples to oranges from year to year?

The inclusion of English learners poses dilemmas

Perhaps the biggest point of contention among state officials concerns the inclusion of EL test scores in the API. The official recommendation by the PSAA Advisory Committee was to exclude all test scores of English learners. Those opposed to including EL student scores in the API say that students who take a test in a language they do not understand are randomly guessing. Therefore, any improvement (or decline) in test scores over time is not proof positive of learning. Not only is an index that relies on SAT-9 scores for EL students flawed, opponents say, but it also sends misleading signals to parents of EL students about the relative progress of their children's schools.

However, the governor and SBE ruled against the committee's recommendation, arguing that the API must include EL student scores on the SAT-9. From their perspective, some accountability mechanism must be in place for the more than 1.4 million EL students who comprise 25% of the state's overall student population. Otherwise, schools might ignore the learning needs of such students. While opponents of including EL scores agree that accountability must exist for EL students, they say the focus should be on creating an accurate system, not relying on a faulty one.

Conclusion: Success of API depends on many factors

The success of the API will undoubtedly rely on many factors. At the top of the list is including multiple measures of performance to derive the API—rather than relying on a single indicator. In particular, many are calling for more time and resources to complete performance-based measures that are closely tied to the state's content standards (e.g., the augmented portion of the SAT-9).

How clearly the design and intent of the API are communicated to those most affected—principals, teachers, students, and parents—will also be critical to its implementation and widespread impact. Whether the API, as it evolves, is viewed as a legitimate and fair instrument for measuring school performance by both the policy-making and education communities will also determine the degree to which it remains a viable stimulus for reform.

Finally, policy makers and educators inevitably will want to know whether the cost of creating and using such a complex index system was worth the benefits. As it stands right now, the API is a critical piece in an overall system that has almost $200 million attached to it in 1999–2000. Will this investment in a high stakes reward-and-punishment mechanism steer classroom learning in the desired direction?

Hopefully, the long-term evaluation called for in the PSAA will be helpful in answering these and other questions. **\\E**

API Information Released for Each School in January 2000

✔ Schoolwide API
- 1999 Percent of students tested
- 1999 API score (Scale 200-1,000)
- 1999 API statewide ranks (Scale 1-10)
- 1999 API similar schools ranks (Scale 1-10)

✔ Schoolwide Growth
- Growth target for 1999
- API target for 2000

✔ Ethnic /Racial and Socioeconomic Subgroups
- Number tested in each subgroup
- Subgroups considered numerically significant
- 1999 API for each numerically significant subgroup
- 1999 Growth target for each numerically significant subgroup
- 2000 API target for each numerically significant subgroup

To find all public information regarding a school's API, see www.ed-data.k12.ca.us or www.cde.ca.gov/psaa.

EdSource 1/00

New Study Investigates Accountability Issues

The evolution of accountability in California and the thorny issues raised by the passage of the Public Schools Accountablity Act (PSAA) were explored in a recent study that was completed in November by researchers at WestEd and Management Analysis and Planning (MAP) under contract with the California Department of Education. The study surveyed 200 districts and conducted in-depth interviews in several districts in the spring of 1999. State officials were also interviewed.

Many district and school administrators embrace the notion of accountability, according to the study, but they question the coherence and long-term effectiveness of new state policies. Of special concern is the status of the current state assessment system, which, as yet, does not completely align with locally developed or state standards. Also at issue is whether the PSAA will eclipse pre-existing local assessments and other reform efforts already aimed at improving schools.

Ten recommendations are offered by the researchers, such as the need for the speedy inclusion of standards-based items in the new API.

The CDE-sponsored study, *Evaluation of California's Standards-Based Accountability System: Final Report* (November 1999), will be available in February at http://www.cde.ca.gov/ope/sse and http://www.wested.org/wested/news.html.

This report has been prepared by

Lisa Carlos, writer/researcher

Barbara Miller, researcher

Susan Frey, editor